research

BEST PRACTICES REPORT

Q3 2019

Driving Digital Transformation Using AI and Machine Learning

By Fern Halper

Co-sponsored by

ANSWERROCKET

tdwi

Transforming Data With Intelligence™

research

BEST PRACTICES REPORT

Q3 2019

Driving Digital Transformation Using AI and Machine Learning

By Fern Halper

Table of Contents

About the Author

FERN HALPER, Ph.D., is vice president and senior director of TDWI Research for advanced analytics. She is well known in the analytics community having been published hundreds of times on data mining and information technology over the past 20 years. Halper is also coauthor of several Dummies books on cloud computing and big data. She focuses on advanced analytics, including predictive analytics, text and social media analysis, machine learning, AI, cognitive computing and big data analytics approaches. She has been a partner at industry analyst firm Hurwitz & Associates and a lead data analyst for Bell Labs. Her Ph.D. is from Texas A&M University. You can reach her by email (fhalper@tdwi.org), on Twitter (twitter.com/fhalper), and on LinkedIn (linkedin.com/in/fbhalper).

About TDWI

TDWI, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and research in the business intelligence and data warehousing industry. TDWI is dedicated to educating business and information technology professionals about the best practices, strategies, techniques, and tools required to successfully design, build, maintain, and enhance data management and analytics. TDWI also fosters the advancement of data management and analytics research and contributes to knowledge transfer and the professional development of its members. TDWI offers a worldwide membership program, five major educational conferences, topical educational seminars, role-based training, onsite courses, certification, solution provider partnerships, an awards program for best practices, live webinars, resource-filled publications, an in-depth research program, and a comprehensive website: tdwi.org.

About the TDWI Best Practices Reports Series

This series is designed to educate technical and business professionals about new business intelligence, analytics, AI, and data management technologies, concepts, or approaches that address a significant problem or issue. Research is conducted via interviews with industry experts and leading-edge user companies and is supplemented by surveys of business and IT professionals. To support the program, TDWI seeks vendors that collectively wish to evangelize a new approach to solving problems or an emerging business and technology discipline. By banding together, sponsors can validate a new market niche and educate organizations about alternative solutions to critical problems or issues. To suggest a topic that meets these requirements, please contact TDWI senior research directors Fern Halper (fhalper@tdwi.org), Philip Russom (prussom@tdwi.org), or David Stodder (dstodder@tdwi.org).

Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, especially those who agreed to our requests for phone interviews. Second, our report sponsors, who diligently reviewed outlines, survey questions, and report drafts. Finally, we would like to recognize TDWI's production team: James Powell, Peter Considine, Lindsay Stares, and Michael Boyda.

Sponsors

Alation, AnswerRocket, Hitachi Vantara, Infoworks, Melissa Data, and TIBCO sponsored the research and writing of this report.

Research Methodology and Demographics

Report Purpose. There is much excitement about using artificial intelligence (AI) to change how organizations do business. AI technologies are driving business value in a number of ways. These advanced algorithms are being used to gain deep analytics insights; they are being embedded in software across the analytics life cycle; and they are utilized in business and consumer apps. The purpose of this report is to explore what is occurring in the market and provide best practices for success.

Terminology. AI means different things to different people. This report uses a narrow definition: "AI is the theory and practice of building computer systems able to perform tasks that normally require human intelligence." Here, AI includes machine learning (ML) and natural language processing (NLP) as well as other technologies. ML is a focus in this report because it is more widely deployed compared to other techniques.

Survey Methodology. In June 2019, TDWI sent an invitation via email to the analytics and data professionals in our database, asking them to complete an online survey. The invitation was also posted online and in publications from TDWI and other firms. The survey collected responses from 379 respondents: 21% have been making use of AI for at least two years, 28% have just started to make use of the technologies, 40% were exploring the technologies, and 11% were not using or exploring the technologies. Although not every respondent completed the entire survey, all responses are valuable and are included in this report's data sample.

Research Methods. In addition to the survey, TDWI conducted telephone interviews with technical users, business sponsors, and analytics experts. TDWI also received briefings from vendors that offer products and services related to these technologies.

Survey Demographics. Respondents act in a variety of roles. The majority of survey respondents are directly involved in analytics (35%), followed by those in architecture and development (17%) and IT (16%).

The consulting (17%), telecom/internet (13%), and financial services (11%) industries dominate the respondent population, followed by healthcare/pharma (8%), and government (8%). Most survey respondents reside in the U.S. (53%), Europe (13%), or Canada (8%). Respondents come from enterprises of all sizes.

Position

Business analyst/data scientist	22%
Architect/developer	17%
Corporate IT	16%
Corporate analytics leader	13%
Corporate executive	11%
Consultant	9%
Other	12%

Industry

Consulting/professional services	17%
Telecom/tech/internet	13%
Financial services	11%
Healthcare/pharma	8%
Government	8%
Manufacturing	6%
Education	6%
Utilities/energy	4%
Other	27%

("Other" consists of multiple industries, each represented by less than 3% of respondents.)

Geography

United States	53%
Europe	13%
Canada	8%
Mexico, Central/South America	8%
Asia	7%
Other	6%
Middle East	2%
Australia/New Zealand	1%
Africa	2%

Number of Employees

10,000 or more	30%
1,000 to 9,999	27%
100 to 999	25%
Fewer than 100	16%
Don't know	2%

Company Size by Revenue

Less than \$100 million	48%
\$100–\$499 million	12%
\$500 million–\$999 million	10%
More than \$1 billion	38%

Based on 379 survey respondents.

Executive Summary

There is considerable excitement about AI technologies, including machine learning and natural language processing. Organizations are embracing these technologies to gain better insights, make better decisions, and improve competitive advantage. In fact, AI is at the heart of the digital revolution in analytics occurring today. AI promises to help organizations improve their operations and processes and to drive new revenue opportunities.

Organizations are making use of AI technologies in numerous ways. Some of these may sound familiar, such as using AI to build churn models or predict fraud. Others seem more revolutionary, such as using AI to diagnose cancer or improve crop yield. AI is being used across the organization and across industries. Those organizations that are already using AI technologies are gaining value from it.

Machine learning dominates the technologies in use; over 90% of respondents who are active users of AI make use of machine learning. Many are building machine learning models and putting them into production. Others are building applications. Organizations are also making use of natural language processing for mining text as well as for servicing customers. For instance, chatbots are popular for customer service. Organizations are using deep learning for image recognition. The use cases are wide and varied.

Organizations utilizing AI technologies are doing so primarily by employing skilled data scientists and other team members, such as DevOps. In fact, 67% of organizations deploying AI technologies today state that AI projects are built by data scientists and are deployed into production by DevOps teams. There is also movement to use augmented intelligence applications, e.g., those where AI is infused into the software to automate functionality such as data cleansing, deriving insights, or building predictive models. Where less than one-third of respondents use these tools today, an additional 50% are planning to use these tools in the next one to two years.

However, employing data scientists or using augmented intelligence isn't enough to create a successful AI deployment. AI requires a modern data infrastructure to support new data types and often massive amounts of data. Many organizations are moving to the cloud for data management. They are making use of data engineers and newer pipeline tools to help integrate data and make sure it is trustworthy. They are hiring DevOps teams to deploy models and monitor them in production. They are evangelizing to build excitement and trust.

This TDWI Best Practices Report examines how organizations using AI are making it work. It looks at how those exploring the technology are planning to implement it. Finally, it offers recommendations and best practices for successfully implementing AI in organizations.

Introduction: AI is Here!

Artificial intelligence (AI) is the phrase *du jour* and market hype around it is soaring. Across industries, AI is capturing people's imaginations. Use cases include self-driving cars and smart city applications that can detect traffic jams and reroute drivers, or even help them locate empty parking spots. In medicine, AI is being explored for use in prosthetic devices that can distinguish images and detect pressure; deep learning is being tested on images to identify cancers or whether a tumor is responding to treatment. Agricultural advances include ground robots and in-field monitoring using drones that can determine whether a crop needs watering or is being infiltrated by weeds.[1] In space exploration, deep learning is used to classify radio-telescope signals in the search for extra-terrestrial intelligence (SETI).[2] AI is even being used to generate music!

New AI use cases are capturing people's imaginations. In agriculture, AI is used in drones to determine whether a crop is being infiltrated by weeds.

AI and Digital Transformation

Yes, there is excitement about AI, but what is AI, exactly? The idea that machines can act intelligently has been around since the ancient Greeks, yet there has been no real consensus about what the term *artificial intelligence* actually means. Back in the 1950s, when John McCarthy used the term, he described it as "making a machine behave in ways that would be called intelligent if a human were so behaving." There has been debate about the term ever since. Researchers will describe general AI as more similar to human intelligence in that it can solve "various problems which were not assumed in the design phase."[3] General AI can be applied to broad tasks and can adapt and learn, similar to the human brain. However, most AI in use today is narrow in scope and is designed to solve a specific problem. Examples of narrow AI include speech or image recognition, or a question-and-answer system.

From a technology perspective, AI is an umbrella term that includes numerous methodologies and techniques (see sidebar on the following page). AI makes use of technologies in the fields of mathematics, computer science, computational linguistics, cognitive sciences, and robotics, to name a few. These technologies are being used against disparate and often complex data to support a host of use cases. Two of the most popular AI technologies are machine learning and natural language processing. In fact, when many vendors use the term *AI*, they are typically referring to one or both of these technologies.

AI is also at the heart of the digital transformation that is occurring across many businesses today. Organizations see AI as a path to help them gain competitive advantage. In fact, in this survey, 90% of respondents agreed with the statement that "advanced technologies such as AI are critical for competitive advantage." Whereas the digital transformation that occurred in the 1990s and early 2000s was about the internet and e-commerce, this digital transformation is about utilizing advanced analytics, such as AI technologies, to change how businesses operate and how they provide value. This includes improving operations, better servicing customers and providing them with better experiences, building new applications to monetize AI, gaining deeper insights for innovation, and much more.

In this survey, 90% of respondents believe that AI is critical for competitive advantage.

[1] See what John Deere is doing, for example: https://www.deere.com/en/our-company/technology-and-innovation/

[2] Mark Williamson, "Space: the new AI frontier?" http://eandt.theiet.org/content/articles/2019/02/space-the-new-AI-frontier

[3] See this interview with Dr. Hiroshi Yamakawa, https://futureoflife.org/2017/10/23/understanding-agi-an-interview-with-hiroshi-yamakawa/

The Many Use Cases for AI

Although futuristic use cases of AI have been in the news, many organizations are gaining value from using AI technologies to solve more familiar business problems. For instance, when we asked survey respondents to tell us about use cases utilizing AI in their own organizations, they noted fraud detection, churn analysis, image classification and analysis, failure detection, plant maintenance, anomaly detection, and robotic process automation. In other words, organizations are reaping the benefits of AI technologies in ways that aren't necessarily futuristic but are significant and potentially transformative in improving operational processes and customer experience.

In this report, the group currently using the technology is referred to as the "active" group and those who are planning to use the technology are referred to as the "investigating" group. Remember that 49% of respondents are already using some kind of AI technology; of this 49%, almost half (21%) have been using it for more than two years. Forty percent are investigating the technology now. Eleven percent are not using the technology. Note, however, that these numbers should not be viewed as an adoption rate because respondents tend to gravitate to surveys that they can relate to.

A GLOSSARY OF POPULAR AI TECHNOLOGIES

AI is an umbrella term that includes a variety of technologies. Popular AI technologies include:

Machine learning (ML): Systems learn from data to identify patterns with minimal human intervention. Machine learning originated in the field of computer science. Popular machine learning algorithms include decision trees, neural networks, and naïve Bayes classification.

Deep learning: A subfield of machine learning where algorithms learn functions that can classify complex patterns such as images; often uses deep neural networks.

Transfer learning: A machine learning method where a model used for one problem is applied to a new problem.

Natural language processing (NLP): Systems that can read, analyze, and understand human language with the goal of simplifying human/computer interaction. Many rely on machine learning algorithms, although other techniques, such as rule-based approaches, are used. A popular example is text analytics to extract entities, concepts, and sentiments from text data. NLP also includes speech recognition.

Natural language generation (NLG): Systems that can automatically generate a narrative from data. Considered a type of NLP/computational linguistics.

Speech recognition: Another subfield of computational linguistics that enables computers to recognize and translate spoken language into text (e.g., speech to text).

Computer vision: A field of computer science where computers can obtain information from images or other multidimensional data.

Computer reasoning: A field of computer science helping to enable systems that can reason as humans do.

The most popular use case for AI is building predictive models using tools such as machine learning.

We asked respondents what use cases dominate when they think about AI in their organization. As illustrated in Figure 1, there are a range of use cases cited as top of mind for AI among both active and investigating groups. The most popular, by far, is building predictive analytics models using tools such as machine learning (85%). These models are often embedded in systems in order to take action. However, other (and sometimes overlapping) use cases cited by respondents include:

- **Automating the generation of insights using tools with augmented intelligence.** Augmented intelligence refers to software that has internal "smarts"; often machine learning is infused into the product. For instance, vendors are providing tools that can automate the building of predictive models to generate insights. Some vendors provide natural language interfaces for users to ask questions. Others provide automatic insights from the data. Forty-seven percent of respondents cited automating the generation of insights using tools with augmented intelligence as a use case for AI. Augmented intelligence was also cited as an important use case for improving data quality by 32% of respondents. Augmented intelligence is discussed in more detail later in this report.
- **Building intelligent apps.** An up-and-coming area of interest for AI is embedding machine learning models and other AI technologies in applications that require intelligence. Building intelligent apps for uses within (38%) and outside (27%) the organization was cited as an important use case for AI. AI apps are everywhere. There are smartphone apps that suggest words or phrases to users; they process speech or detect junk photos. There are vertical AI solutions for industries including education (e.g., identifying gaps in student knowledge, developing customized learning programs) and energy and utilities (e.g., smart meters, drones inspecting wind turbines).
- **Robotic process automation (RPA).** The goal of RPA is to automate manual processes through the creation of software robots to replace or supplement repetitive tasks performed by humans. Typical examples of RPA include processes that operate 24/7, such as complex order processing or payroll processing. RPA is evolving to enable AI and machine learning technologies to help handle high-volume, repeatable tasks with more accuracy. For instance, machine learning can be used to identify images or defects as part of a process. In this survey, 31% of respondents thought of this use case for AI.
- **Conversational AI.** Conversational AI is a form of interaction where the computer converses with the human. It includes speech-based assistants such as Alexa, chatbots, and messaging apps. About 15% of respondents selected this as a top AI use case for their organization.

When you think about AI in your organization, what use cases dominate?

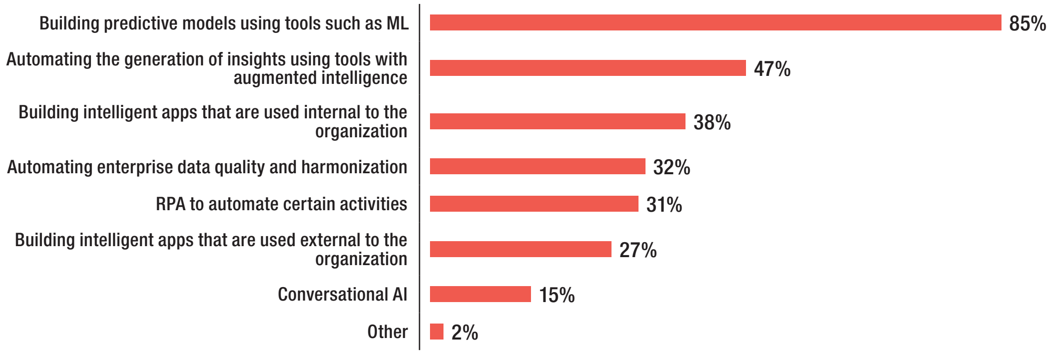

Figure 1. *Based on 294 respondents in the active and investigating groups. Maximum of three responses allowed.*

The State of AI

To further understand the *current* state of AI, we asked respondents about their experiences and plans with AI. For those who are not using the technologies and have no plans to do so, we asked, "why not?"

Machine Learning Dominates AI Today

We asked the active group what kind of AI technologies they currently use. Not surprisingly, machine learning dominates the results.

TDWI research indicates that the vast majority of organizations that deploy AI solutions use machine learning as part of that solution.

Machine learning rules. The vast majority (92%, Figure 2) of active group respondents use machine learning. This is where we see organizations typically start with AI. Predictive analytics against structured data is in mainstream adoption; it is a natural step for organizations to extend what they are doing to incorporate machine learning algorithms. Once organizations get comfortable with machine learning, they may also use deep learning for use cases that often involve image recognition. Deep learning was used by 41% of respondents. Computer vision, which uses machine learning, was in use by 24% of respondents. Many active group members are using machine learning in models they are putting into production (64%, not shown). Others are developing applications that include machine learning (43%, not shown).

Sixty-four percent of active group respondents are already using some kind of NLP (including text analytics).

NLP (including analyzing text) is important. In this survey, 64% of active group respondents were already using some kind of NLP, including text analytics.[4] Text analytics has been used for years by forward-looking companies to extract entities (people, places, things), concepts (words and phrases that indicate a particular idea), themes (groups of co-occurring concepts), and sentiments (positive, negative, neutral) from text data. It has been used in social media analysis and analysis of claims notes, among other text documents. This extracted data is also used as attributes in machine learning models. For instance, an organization might extract information (such as sentiment) about various entities (such as customers) that could be used as input when building a churn model. This extracted text can be highly predictive.

At TDWI, we've been monitoring how organizations are using unstructured data, particularly text data, for quite some time. Some companies have been relatively quick to utilize deep learning for projects such as image classification, but the notion of analyzing/utilizing text in applications has not received as much attention. We typically see that overall, text analytics is moving into mainstream adoption. In fact, 43% of active group respondents (not shown) were already analyzing text to better understand customers. That number could double in the next one to two years if users stick to their plans. That is good news because text can provide great value and insight.

NLP (using natural language processing and understanding) is also a strong use case.

- **Chatbots.** Chatbots are a popular use case for NLP. In the active group, 35% of respondents (not shown) are developing bots for customer service. Newer chatbots have a better understanding of language and are more interactive. Some organizations use chatbots to help answer routine questions. Some chatbots help route calls to the appropriate person. Personal assistants such as Siri may use speech recognition to convert speech to text and then use NLP to understand the intent of what is being said. The assistant can then act on that, perhaps using NLG to respond.

[4] Machine learning is often used against text data to help in natural language processing. The system can be trained to understand a combination of words or characters.

- **Infused in applications.** NLP technologies can be infused into analytics applications. For instance, NLP is infused into BI systems that can suggest insights using a natural human-based language interface. Another AI technology, machine reasoning (which enables a computer to make and interpret connections) is used by some vendors to improve data quality. Machine reasoning applications reason from ontologies—formal specifications of the terms in the domain and relations among them.[5] In that way, an automated data quality tool could use machine reasoning to match names based on certain interpreted patterns.

What kind of AI technologies do you currently use? Please select all that apply.

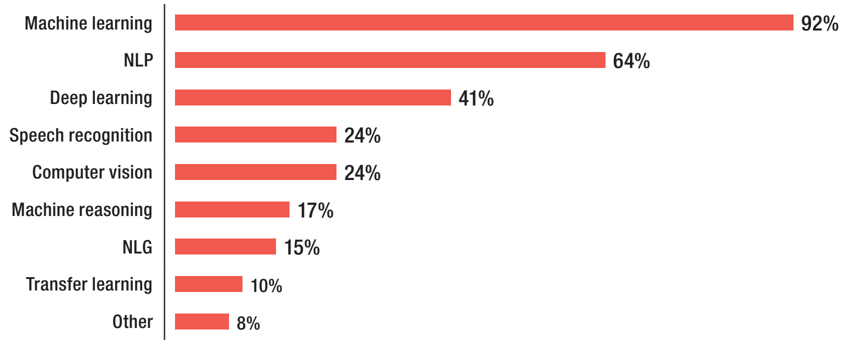

Figure 2. *Based on 157 respondents in the active group. Multiple responses were allowed.*

Deployments and Departments

Though certain AI technologies such as machine learning are moving into mainstream adoption, we wanted to get a sense of how many deployments the active group currently had in place, as well as what departments were utilizing AI. Sometimes it appears that organizations get stuck in the pilot phase and can't move forward. We wanted to determine whether the active group programs were mostly in the pilot phase or whether the group had more than one AI project deployed. We also wanted to determine how advanced the active group is with regard to AI deployments.

Most active groups have multiple deployments. About 25% of active group respondents either had pilots or only one project in production; most have more. For instance, 23% of respondents had two to four projects in place. Another third had five or more projects deployed. In other words, AI is being deployed and is in production in multiple projects in the active group. (Figures not shown.)

The majority of those deploying AI projects had more than two projects already in place.

AI projects are spread among departments. We also wanted to understand where in the organization AI was being used both currently and in the next few years. As Figure 3 illustrates, Operations and IT are two top areas where projects are deployed for the active group. This includes use cases such as predictive maintenance. However, AI use is spread throughout the organization. For instance, 38% currently use it for customer service (e.g., chatbots, next best offer); 35% use it for marketing (e.g., customer behavior analysis, churn). It is even used in HR (e.g., employee churn, employee hiring).

[5] Gruber, T.R. (1993). "A Translation Approach to Portable Ontology Specification." *Knowledge Acquisition* 5:199-220.

Adapting to new contexts and objectives. One of the holy grails of AI is the ability for machines to adapt to new contexts and objectives without human oversight. In other words, these are AI systems that can learn and adapt autonomously. The smart-driving car is one example of AI that would utilize data (from cameras and sensors, for example) in real time and then adapt. Automated assembly lines are another example. However, there are also more traditional use cases such as recommendation engines adapting to new buying behavior or risk management solutions that can adapt to changing fraud patterns. Only 13% (not shown) of active group respondents have projects in the field today which do this. Another third are working on it. This illustrates that although organizations are utilizing AI technologies, some aspects of AI are still in their infancy.

AI is not real-time, yet. Likewise, when asked whether their organizations have an architecture that utilizes real-time data, 49% of the respondents stated that they don't support real-time data now (not shown). However, for the active group, only 28% said that they didn't use real-time data now. In fact, 45% of the active group that has been using the AI for more than two years makes use of streaming technology compared to 29% overall (not shown). This, no doubt, makes sense for some of their use cases.

In what departments is AI being used in your business today? In the next few years?

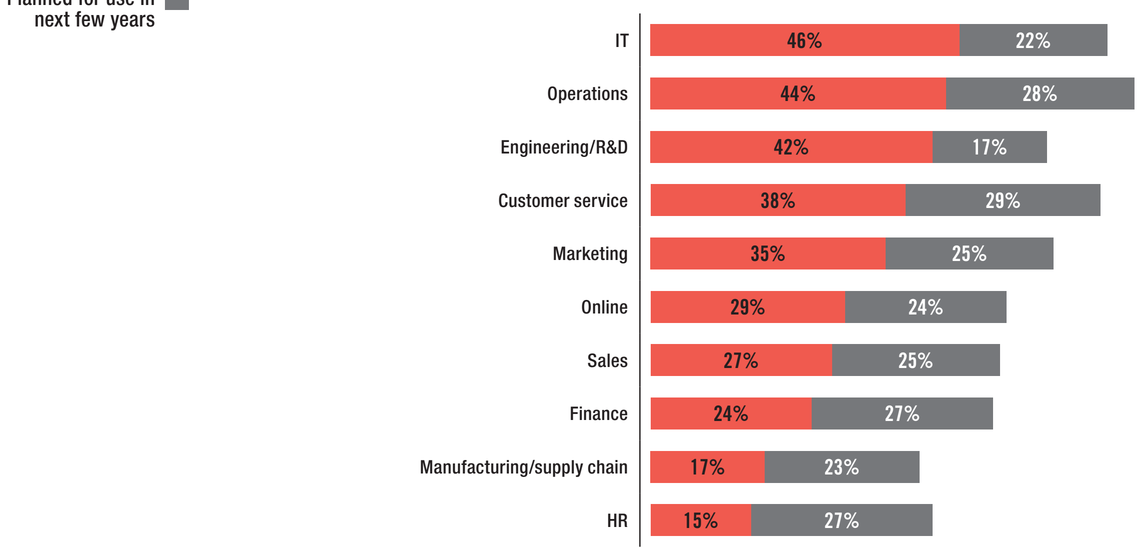

Figure 3. *Based on 157 respondents in the active group.*

Some industries appear to be more advanced than others. Unsurprisingly, industries such as telecommunications, internet, and high-tech appear to be more likely to have been using AI technologies for more than two years than industries such as education. Likewise, larger companies (over $1B in revenue) seem to be more likely to have been using AI technologies for more than two years than have smaller companies.

Augmented Intelligence

As mentioned above, vendors are putting significant effort into software that can assist or augment human intelligence. The idea behind augmented intelligence isn't to replace humans but to help them with tasks. Solutions that bill themselves as augmented intelligence include automation, but typically the system is not completely automated.

The use of augmented intelligence will be a big growth area if users stick to their plans.

In analytics, AI technologies are being infused across the analytics life cycle, including in data preparation, generating analytics insights, and in building analytics models. The use of augmented intelligence (also called *"smart" tooling*, *AutoML*, *intelligence augmentation*, and *intelligent experience*) will be a big growth area if users stick to their plans (Figure 4). Some areas of growth include:

- **Data preparation.** Data preparation covers a range of processes that begin during an organization's initial ingestion of raw structured and unstructured data from one or multiple sources. Data preparation processes focus on determining what the data is and improving its quality and completeness, standardizing how it is defined and structured, collecting and consolidating it, and transforming it to make it useful, particularly for analysis. Often, this is a time-consuming, manual process.

  AI-infused tools on the market embed machine learning algorithms into the software for data preparation. For example, machine learning is being used to assess data quality and accuracy (e.g., address mapping). Automation is used to transform data for analysis, such as automatically creating common ratios from attributes in the data file or for use in analysis. In this survey, 14% are already using AI-infused data quality tools that automate the cleansing of data using techniques such as machine learning or machine reasoning. Another 58% plan to use the software in the next one to two years (see Figure 4).

- **Augmented insights.** Vendors are also providing tools that can derive insights from data and present those insights to the end user. Some of these tools use a natural language interface to search and analyze data to find insights. The user asks a question in a natural way and the software returns potentially relevant answers. BI and analytics solutions can also employ machine learning to explore data automatically and spot trends and patterns that users working with standard querying and reporting capabilities may not have seen. These are then presented to the user. In this survey, 15% of respondents are currently using these kinds of tools. Another 50% plan to do so in the next one to two years (see Figure 4).

- **Automated model building.** Data scientists and statisticians are often in short supply. Many vendors are offering tools that help business analysts and even business users construct machine learning models. In some of the tools, all the user needs to do is specify the outcome or target variable of interest, along with the attributes believed to be predictive. The software picks the best model. Other tools are even more automated. In this survey, 23% of respondents are currently using this kind of software. Another 53% plan to use the software in the next one to two years (see Figure 4).

The active group was more likely to be using all of these tools today in some capacity, compared with the investigating group.

These tools can be a big productivity booster. Some benefits that respondents noted include quick deployment of pilots, as well as ease of use for business analysts and those with knowledge of the business. Respondents also noted that these tools can help save time and increase productivity, and that automation of certain tasks can allow people to focus on other areas and even free users from boring tasks. Some respondents stated that these tools would help reduce errors and improve data quality.

Augmented intelligence tools can help save time and increase productivity.

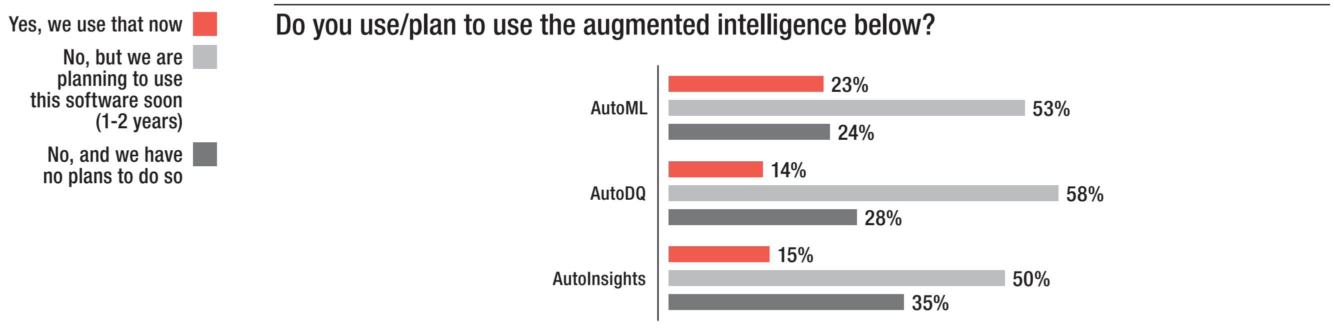

Figure 4. *Based on 272 respondents from both the active and investigating groups*

Some respondents felt that, in the wrong hands, the tools could lead to bias without the user even noticing.

Survey respondents noted concerns about these tools. The primary area of concern was related to tools that produce automatic insights or models and the skills of those using them (as opposed to data preparation/quality tools). Some respondents were concerned that the tools were "black boxes" and felt that, in the wrong hands, the tools could lead to biased results without the user even noticing. Additionally, many respondents felt that users needed to be able to explain key contributing variables and interpret the result of augmented insights, and that those who use the tools may not have the skills to do that, even if the tool provided some of the necessary information. Others were concerned about the loss of jobs associated with automating insight delivery. Some were still not familiar enough with these kinds of tools to have an opinion.

TDWI has written about these tools.[6] We've recommended that users have the skills to verify the insights produced by these tools. For augmented data modeling tools (e.g., AutoML), organizations should choose tools that are transparent so users know what is happening behind the scenes. Additionally, users should understand the techniques used to automate model building and should be able to interpret the results themselves. Enterprises must have a process in place to validate the models before they are put into production.

Open Source for AI

Previous TDWI research has indicated that open source tools such as R and Python are important for AI.[7] These projects contain libraries for many AI techniques and methodologies. For example, R contains many packages for machine learning and NLP/text analytics. Python also contains many libraries to support AI.

Active group respondents are big believers in open source tools such as R and Python. For example, when asked to rate the importance to the AI effort of open source analytics tooling, over 50% of the active group rated tools such as R and Python as extremely important (not shown). Vendors that provide proprietary solutions have bought into the fact that open source is likely here to stay. They are providing users with tools to insert models and other analytics built in open source into their solutions.

Open source can be extremely valuable for building AI applications because of its community of innovation. However, newer users should be aware that utilizing these tools requires specific skills. Additionally, many open source tools do not provide an easy way to address some of the backend aspects of advanced analytics, such as managing and monitoring models.

[6] For more information, see the 2018 *TDWI Best Practices Report: Practical Predictive Analytics* and the 2018 *TDWI Best Practices Report: BI in the Age of Big Data and AI*, online at tdwi.org/bpreports.

[7] For more information, see the 2018 *TDWI Brest Practices Report: Practical Predictive Analytics*, online at tdwi.org/bpreports

Why Not Use AI?

We stated in the invitation for the survey that we wanted input from everyone, and a few respondents (11%) reported not using AI. (As stated, these percentages should not be viewed as an adoption rate.)

For those who do not use AI-related technologies, the top three reasons were that the organization is still focused on basic analytics, they don't have the skill for AI, or they are still dealing with the data infrastructure to support analytics and AI. Only 8% (not shown) stated that they don't need the functionality. When asked about whether they think AI is valuable, the majority (65%, not shown) stated that they wished they could make use of it in their organization.

The majority of those organizations not using AI stated that they wished they could make use of it in their organization.

Infrastructure to Support AI

AI often makes use of high volumes of data from disparate sources. This includes traditional structured data as well as text data, image data, machine data, and other types. To build ML models or develop NLP apps, organizations will also need the infrastructure to support managing multiple data sources as well as the skills and tooling to build potentially complex data pipelines to make sure they can pull disparate data together and trust that data.

We asked respondents what kinds of data they are using for analytics now and what they planned to use in the next few years. As TDWI has seen in other research, the vast majority of respondents (91%) are using structured data for analysis (Figure 5). However, for AI, other data is being used as well. For instance, 64% of respondents across both the active and investigating groups are analyzing text data. The percentage is higher for the active group (85% of those using AI for more than two years, not shown). Additionally, organizations are managing larger amounts of data. The majority of respondents are managing terabytes of data and about 8% are managing petabytes (not shown).

What kinds of data do you use now? Within the next 3 years?

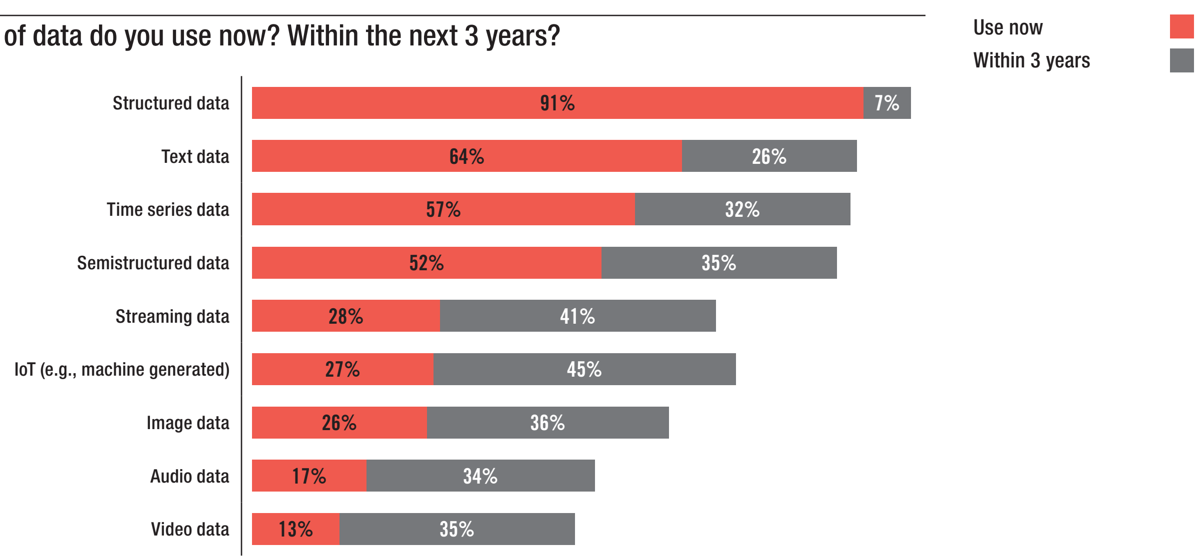

Figure 5. *Based on 258 respondents in both the active and investigating groups.*

Multiplatform Data Environments Are Important

In order to deal with new data types and complexities, many organizations are deploying a multiplatform data environment.

To deal with new, more complex data, many organizations are deploying what TDWI refers to as "multiplatform data environments." These environments typically include traditional relational database management systems along with newer DBMSs, often based on cloud systems; Hadoop and its ecosystem; streaming platforms; and columnar databases. Organizations make the move to multiplatform environments because they need particular platforms to support certain workloads. For example, an organization might generate much of its data in the cloud. Rather than move it on premises, it makes sense to keep that data in the cloud, especially if it will be analyzed there. Alternately, an organization might not want to store text or image data in an on-premises data warehouse.

Figure 6 illustrates the kinds of platforms the active group is using for data management compared with the investigating group.

- **Both groups use the data warehouse.** Similar to previous TDWI survey results, the vast majority (over 75%) of both the active and the investigating groups are already using a data warehouse on premises to support their analytics/AI efforts. However, even in the investigating group, more than half already have another platform in place or plan to have another platform in place within the next year or two. In other words, utilizing a data warehouse alone is not typically the way forward for AI, although some organizations choose to do this.

- **The active group is more likely to use a modern platform.** Previous TDWI research has shown that as organizations mature in their analytics, they often utilize a greater number of platforms. In other words, they move to a multiplatform environment. This same pattern emerged in this survey. In fact, with the exception of the data warehouse (either on premises or in the cloud), the active group is at least twice as likely to utilize a modern platform, typically in conjunction with the data warehouse. For instance, a data lake in the cloud is used by 38% of active group respondents and only 12% of the investigating group. A cloud-native data warehouse (e.g., a warehouse specifically architected for the cloud) is used by 21% of the active group and only 9% of the investigating group.

The cloud is a major area of growth for AI data management.

- **The cloud is a growth area.** Across both groups, a top area for growth is the cloud. For instance, 23% of all respondents are using an enterprise data warehouse in the cloud today and an additional 47% plan to deploy one in the next few years. Likewise, 26% are using a cloud data lake today and an additional 47% are planning to use cloud data lakes in the next few years. Although 16% are using cloud-native data warehouses today, 33% more are planning to use cloud native data warehouses in the next few years (all not shown).

Figure 6. *Based on 258 respondents in both the active and investigating groups.*

Data Quality Is Imperative in AI

Data quality is regularly at the top of the list of technical challenges for analytics.

Data quality is critically important for AI success. As the old saying goes, "garbage in, garbage out." For example, machine learning algorithms do not perform well with garbage data.[8] TDWI regularly sees data quality at the top of the list of technical challenges for data management and advanced analytics. Poor data quality includes misunderstood data, duplicate data, incomplete data, out-of-date data, and misspelled or misfielded data. Data quality is also a foundation for data auditability and data governance.

In this survey, we asked active group respondents, "How does your organization ensure that the data quality you use for AI is sufficient?" For those investigating the technology we asked, "How do you currently handle data quality?" (Figure 7).

- **Many respondents profile and cleanse their data.** The majority of respondents perform tasks such as data cleansing to remove duplicates and nonstandard data types (67%). They also profile data to manage compliance with rules (51%). In the active group, these numbers were slightly higher (74% and 56% respectively, not shown). Of course, there is still room for improvement. In this survey, 13% said they don't implement any kind of data quality steps.
- **A small percent of organizations are using AI-infused tools.** As previously mentioned, some vendor solutions are using AI technologies to identify (and often correct) problems in data. For example, some software uses machine learning to determine whether an address written three different ways is the same. Other software employs sophisticated, built-in AI algorithms pre-trained to do the same thing. Still other tools can autocomplete addresses or map addresses to geocodes. In this survey 11% of respondents were using AI-infused tools to improve data quality. (This is consistent with the approximately 14% of respondents who said that they use auto data cleansing tools.) However, as we saw, there is great interest in these kinds of tools.

[8] V. Gudivada, A. Apon, and J. Ding. "Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations," *International Journal on Advances in Software*, 10.1 (2017), pp. 1-20.

Metadata management and data lineage are areas for improvement for organizations.

- **Managing data lineage and metadata are areas for improvement.** We also asked respondents whether they track metadata or manage data lineage. Data lineage is important for data quality because it allows users to investigate data quality issues. Data lineage provides information about data from source to destination; it describes where data originated and how it has been changed and transformed. This might include how the impact of changes to one data element could affect other systems.

In this survey only 32% of respondents tracked data lineage. Somewhat more (48%) managed metadata. Metadata provides descriptions about the data. This includes technical metadata about data types and structures. It includes business metadata that describes data in user-friendly terms. It also includes operational metadata that records access to data by users and applications. A good metadata strategy provides information about the data and can help identify data quality problems (e.g., if formats are incorrect or if data is missing). This, in turn, helps engender trust in the consistency and accuracy of the data, which is critical for analytics.

Current users: How do you ensure that the data quality you use for AI/ML is sufficient? Those exploring: How do you currently handle data quality?

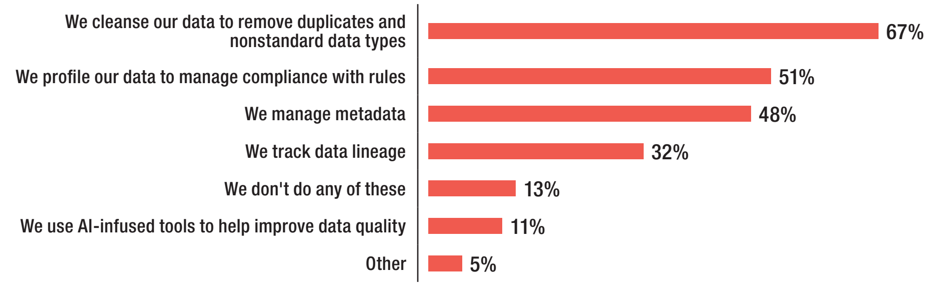

Figure 7. *Based on 256 respondents in the active and investigating groups.*

USER STORY A SOUND DATA INTEGRATION STRATEGY IS KEY FOR MACHINE LEARNING

"With regard to predictive analytics and machine learning, I think everybody is interested in doing this kind of thing," said a senior data warehouse architect at a consulting firm. "The promise of predictive analytics is strong enough for most companies to want to use it. The difficulty is delivering on that. The data doesn't cleanse or organize itself in a way that you can just plug it into a model. You need an engineer and ideally a repeatable curation process in order to use it in that way."

That may involve new tooling to deal with data integration and issues of metadata and master data. According to this architect, "Data integration tools are the new sexy, but the learning curve carries a price. What I see happening—especially at midsize companies—is that they often hire smart people out of school who know some open source tooling. This often results in a dozen homegrown systems that are a horrible mishmash. They try to get them to talk together via APIs instead of taking a centralized approach to curation (and developing data management as a practice). There is a risk to that kind of approach. If you don't have an enterprise data architecture, it adds unnecessary complexity and expense and makes it harder to move into more advanced analytics."

Data Pipelines Become Key

For digital transformation to succeed, an enterprise must harness the data it needs for analytics. With AI, this task more often falls to the data engineer. The data engineer works with data architects and data scientists to help build the often-complex data pipelines for more advanced analytics. Typically, this starts with ingesting data from multiple sources, integrating it, preparing and curating it, and developing features used for input to models. Often, data engineering is a manual process, with data engineers building these pipelines. That can mean that many data engineers need to be hired to assemble and maintain many data pipelines. As one data scientist interviewed for the report put it, "The slowest part of the model-building process is not having enough data engineers."

"The slowest part of the model building process is not having enough data engineers."

Data pipelining tools can accelerate the process. The tooling has attracted renewed interest, in part because of the complexities involved in assembling big data for more advanced analytics. There are a number of vendors that provide data pipelining tools. They typically provide a drag-and-drop interface with connectors to many data sources, whether they reside on premises or in the cloud. In some instances, users can save the workflow, document it, and make it available for reuse. This saves time and money. Some tools can automate the workflows so organizations can schedule and run them. Some tools even automate tasks at every stage. For instance, at the ingestion stage, certain tools can automate change data capture and can merge and synchronize the data. This is extremely important for machine learning models that will require fresh data in the production stage.

Data Catalogs Build Trust

Along with pipelining tools, TDWI is seeing increased interest in modern data catalogs. In multiplatform data environments, data comes from disparate sources, in disparate formats, and from disparate systems—both internal and external to a company. As organizations move to analyze this data, they want to easily access, understand, and trust it. This is where the modern data catalog comes in.

A fully operational data catalog documents large numbers of data sources, targets, and structures. As users look to utilize vast amounts of new data sources, they are more often want to use cataloging software that is AI-enabled (as is the case with data preparation, automated insights, and AutoML software described above). Some modern cataloging tools parse and deduce credible metadata. Other tools scan each new data set for sensitive data and tag it appropriately so that tag-based security can be applied. Some tools automate the cleansing of data and some automatically discover and suggest missing lineage between data sets. In addition to automation using AI, some modern catalogs embed NLP functionality in the data catalog that helps users ask questions of their data in a natural way. Some even make recommendations as a query is being written, similar to autocorrect. In some products, machine learning is used to help translate file names into more understandable terminology (e.g., a file labeled "drg" being translated to "Drug Related Diagnosis").

These next-generation catalogs often also contain crowdsourcing and collaboration features. Users can rate different data sources and provide information about how they used the data and whether it met their needs. Some use machine learning to make recommendations based on usage data. All of this, in turn, helps build trust in the data—that it is accurate, reliable, timely, and reasonable. Trusted data is critical for organizations looking to move to AI.

The Push/Pull: Skills Needed for AI

Fifty-eight percent of respondents cited skills as a top barrier to AI.

Organizational issues are often the top impediment to successfully deploying analytics, and skills are always cited as the top challenge for any advanced analytics endeavor. In this survey, for instance, access to talent to build AI applications was viewed as a top challenge by 58% of respondents; over 20% higher than the other challenges (Figure 8). It can be complex to build applications that utilize machine learning models, NLP, or other AI technologies.

In addition to understanding the math and statistics involved in models, data scientists and app developers need to understand how to interpret model output. Open source skills, such as those for R and Python, are used by many organizations and are constantly in demand. Computational linguistics skills may be needed for organizations building NLP applications. Data engineering and DevOps skills are needed to build pipelines and operationalize models. Previous TDWI research has indicated that many organizations are planning to improve the skills of their business analysts to help build models, often by using augmented intelligence.[9]

What were or will be the biggest barriers to adoption of AI in your company?

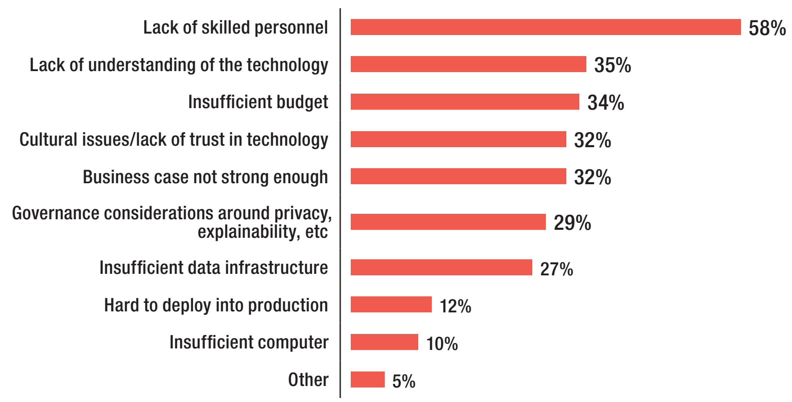

Figure 8. *Based on 268 respondents from both the active and the investigating group. Up to three responses allowed per respondent.*

Data Scientists, DevOps, and Business Analysts

To better understand the dynamics of this push/pull, we compared the active and investigating groups across multiple attributes. These included analytics sophistication, kinds of data used, who is building models and deploying them, and how the active group views advanced analytics.

Not surprisingly, organizations that make use of AI are likely to be more sophisticated in data management and analytics.

- **The active group:** Analytics is widely accepted and adopted at organizations actively utilizing AI. Not surprisingly, these organizations tend to be more sophisticated than those investigating the technology—they are more likely to manage petabytes of data; to be analyzing disparate data sources; and to utilize disparate platforms for data management. The active group is also more likely to view the use of AI as a competitive differentiator and to see the value of AI applications than the investigating group.

  The active group is using AI for a wide range of use cases across the organization, as previously discussed. These AI projects are typically built by data scientists and deployed into production by DevOps—the group that is responsible for operationalizing, managing, and monitoring AI deployments (Figure 9). Fewer than 20% of active group respondents

[9] For more information, see the 2018 *TDWI Brest Practices Report: Practical Predictive Analytics*, online at tdwi.org/bpreports

stated that business analysts are responsible for AI projects. In these organizations, data scientists and app developers are leading the charge.

The DevOps team can be overlooked, but organizations that are successful typically have a DevOps team in place to help validate models and put them and other AI apps into production. DevOps teams are often skilled in newer open source technologies such as Spark and Python, as well as in understanding the output of models built using commercial tools. For example, one DevOps specialist (called an analytics engineer in his organization) interviewed for this report said that although data scientists may use both open source and commercial tools to build ML models, his team has a process and framework in place and will typically convert the model code into an implementation code.

DevOps is also responsible for tracking models in production. Typically this will require some sort of model registry. It is easy to keep track of a few models, but once many models go into production, it is important to register and track them, including metadata about the model (date created, who created it, attributes used, and so on). DevOps also tracks the model once it is in production to look for degradation; they may use algorithms to check for this. If the model starts to degrade, an alert is sent to the team.

DevOps is key to AI success. DevOps puts models into production and keeps track of them.

- **Investigating group.** Although those actively utilizing AI technologies are using data scientists to do so, those exploring the technology are more likely to be looking to business analysts, often in conjunction with data scientists to build models and deploy other forms of AI. Whether this will be a successful path forward remains to be seen. Active group respondents, for instance, are more likely to disagree that users of all skill levels should be able to deploy machine learning models; those investigating the technology are more likely to agree. Additionally, very few respondents cited that using tools that infuse AI into the data and analytics life cycle is a top best practice for getting AI projects off the ground (Figure 11). Yes, organizations will make use of these tools, and some in both groups are doing so already, but the goal is to be careful with the tools. That means having check points at different steps in the process to make sure the model is valid and tested.

The differences between these two groups illustrate some of the push/pull in the market. On the one hand, data science skills are in short supply. Vendors are providing easy-to-use tools to help fill the gap. On the other hand, ML models may be complex and not easy to build. They require the right features, testing, and model tuning. They require understanding how to interpret model output and determining whether the model is good or not. For example, as one data scientist said, “It is too easy to get false results. The model can be 99% accurate, but it may have been built from three data points with one outcome and a million with another. Does the business analyst understand this? Do they also understand concepts such as co-linearity?”

On the one hand, data science skills are in short supply; on the other hand models can be complex and hard to build.

The reality is that organizations will need to hire data scientists for AI; using automated model building alone isn’t sufficient. TDWI has seen that some organizations will make sure that the correct controls are in place so that if a business analyst is using this tooling, a data scientist will check over the model before it is put into production.[10] Stakeholders (such as those who own the systems) and subject matter experts will need to be involved. Data engineers (described in the previous section) will also be needed, although pipelining tools can help to replace the manual drudgery.

[10] For more information, see the 2018 *TDWI Brest Practices Report: Practical Predictive Analytics*, online at tdwi.org/bpreports.

How is AI built and deployed in your organization?

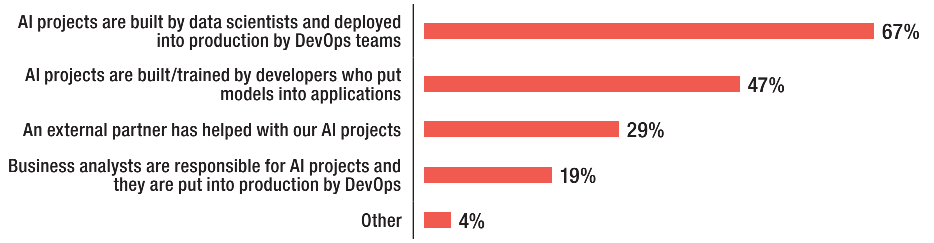

Figure 9. *Based on 157 respondents in the active group. Multiple responses allowed.*

USER STORY ANALYTICS AT SCALE: DATAOPS

According to Mark Teflian, Senior Director, Advanced Analytics at Charter Communications, "We have learned, as we have engineered data, that data is not just a project. It is a product." Charter Communications, a telecommunications and mass media company, deals with huge amounts of streaming data as part of its IP network services. It also utilizes advanced analytics such as machine learning against this data, as part of its converged network strategy. "These are very complex data sets that we merge," said Teflian, "It goes beyond a data extract. It is versioned, with metadata. It is cataloged. It can be watermarked and trace-to-source enabled"

As the company continues to deal with data at scale and apply advanced analytics against the data, the DataOps group becomes increasingly important. DataOps manages the data; data gets updated and enriched. For predictive analytics and machine learning (PAML), that also requires new feature engineering. "If you think about a machine learning model, that model is built against vast amounts of complex data. A group may be putting that into production. Additionally, once it is in production, it is running against ever-changing data. Models will have to be retrained against new data. At the scale that we are dealing with, DataOps—to manage this data—is key. That includes operating the model, and managing it and the feedback loop that is necessary to keep it up-to-date with new data." In the case of Charter Communications, ModelOps (e.g., the group that deploys PAML models) has been merged into the DataOps function.

The team at Charter Communications has multiple skill sets. This includes data scientists, data engineers, ops research scientists, computer scientists, platform architects, agile scrum masters, ModelOps and DataOps. They have integrated the SMEs into these teams, as well. Teflian cautions, "It takes changing a culture and how people work together to drive high-value data and analytics. Think about skill sets. Advanced analytics isn't Money Ball. Opt for a discrete set of skills and then build from there."

Where's the Value?

As TDWI has seen in other research, those respondents who have implemented advanced technologies such as AI tend to be satisfied with their analytics program. In this survey, 39% of the active group stated that its organization has *measured* a top- or bottom-line impact from its AI efforts (Figure 10). Fifty-four percent believe they have obtained value, although they haven't actually measured it. In other words, 93% of the active group believes that AI provides value. Only 7% of the active group has not seen any value from their AI efforts.

Ninety-three percent of the active group believes that AI provides value.

Top- and bottom-line impact comes in many forms. It includes reducing costs by streamlining processes utilizing AI technologies. It includes changing processes, such as intervening when a model predicts that a customer has a high probability of churn rather than waiting for the customer to leave and trying to win them back. It includes top-line revenue from innovative new products and services made possible by AI technologies.

What value have you seen from your AI efforts?

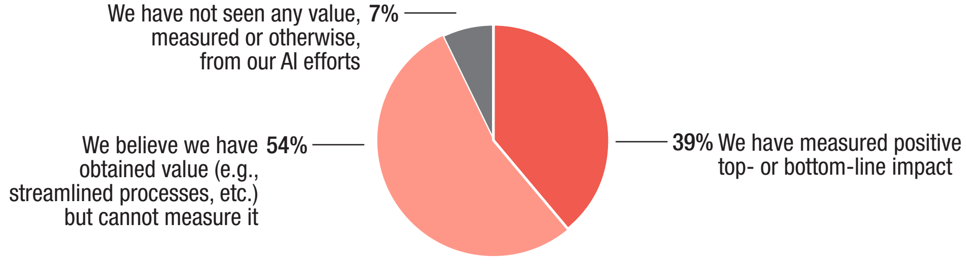

Figure 10. *From 157 respondents in the active group.*

TDWI has referred to the virtuous circle in other reports. As organizations begin to reap the benefits of analytics, they tend to put more advanced analytics in place. The success builds on itself. In fact, in this survey, 61% of respondents stated that AI has improved productivity. More than half (57%) said AI has delivered on its promise. These promises include improved decision making, improved economic outcomes, and competitive advantage (all not shown).

Success Factors

We have already provided some strategies organizations can use to help make their AI projects successful. In the survey for this report, we also asked a series of questions specifically about best practices. As part of the analysis, we compared the active group against the investigating group to look for differences in response. The active group appears to be satisfied with the results of their AI implementations, so it is possible to learn from them. We also looked at those organizations that have measured value to try to tease out best practices further.

- **Start with a business problem.** As Figure 11 illustrates, 53% of respondents stated that starting with a business problem is a best practice for getting AI off the ground. We have seen this in previous TDWI research. Analytics is most successful when it is addressing real business needs. In fact, in a separate question, we asked whether a first AI deployment is/will be driven by a specific business need. Eighty-eight percent of respondents that had measured positive top- or bottom-line impact answered yes. Often these projects were developed around conventional business needs such as customer retention, marketing

Fifty-three percent of respondents stated that starting with a business problem is a best practice for getting AI off the ground.

programs, or supply chain logistics. Yet, these change organizational processes or the way companies do business and are important for digital transformation.

- **Be realistic.** Following on the theme above, 49% of respondents suggested finding basic, achievable projects where AI can help improve time and cost objectives. This was the case across groups.

For those succeeding with AI, the CEO is often the chief evangelist.

- **Get executive sponsorship.** Forty percent of respondents cited this as one of the best practices for getting started with AI. The number rose to 57% among those measuring value with AI and was the *top* best practice (with the two other best practices mentioned in second and third spots). In fact, for those succeeding with AI, the CEO is often the chief evangelist for AI efforts. That may not be the case for all companies, but it shows the importance of evangelizing change.
- **High-quality data is key.** We've discussed this already in this report. It is important to utilize high-quality data for AI or the results won't be trusted. It can take just one problem with the data for an organization not to trust the results of a machine learning model, which can set the organization back in terms of its overall plans. Related to data quality is data governance. The good news is that over 90% of respondents (not shown) said that data governance was either important or extremely important to their AI efforts.
- **Collaborate for results.** AI isn't just the realm of data scientists. As we've discussed, it involves data management professionals, data engineers, data scientists, business analysts, DevOps, SMEs, and more. In this survey, 34% of respondents cited collaboration between business and IT as key for success. This isn't just about working together to get the job done. It is also about building trust among groups, which is critical for digital transformation and change.

What do you believe are best practices to get AI projects off the ground?

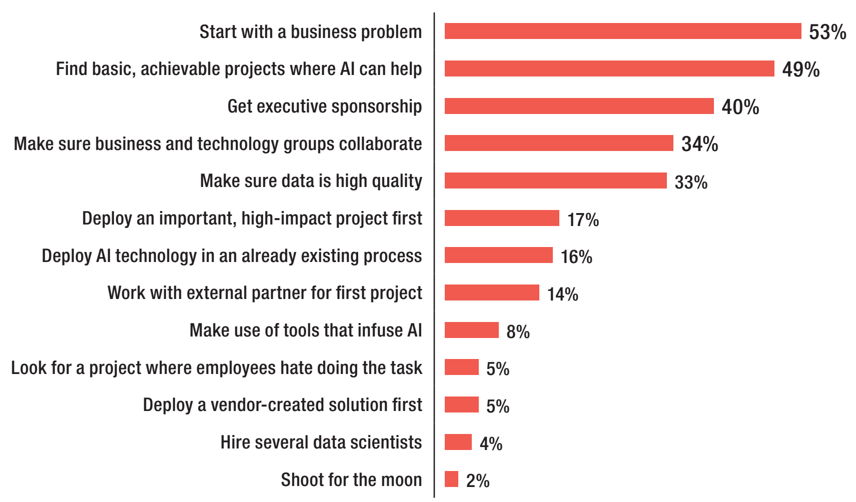

Figure 11. *Based on 256 respondents in both the active and investigating groups.*

USER STORY MAKING THE JOURNEY TO ADVANCED ANALYTICS AT OWENS CORNING

Founded in 1938, Owens Corning is a leader in insulation, roofing, and fiberglass composite materials. About 10 years ago, the company faced a problem that many companies face: the need to define financial metrics and KPIs across the company to enable a single version of truth for reporting. The CFO convened a working group to solve the problem. This working group ultimately evolved to become an analytics center of excellence (CoE), located in IT, to sustain the company's analytics effort. Since then, the CoE increased in size and scope, enabling multiple functions in the company (such as finance, supply chain, sourcing, marketing, and manufacturing) to leverage analytics to drive growth and productivity.

According to Paul Burns, a leader in the analytics CoE, "My role is to work with the business and identify opportunities to leverage data and turn it into an asset. I'm focused on moving organizations up the analytics maturity curve." This means overseeing the data infrastructure and consulting with the business on analytics, including machine learning models. For instance, the data science team has built and operationalized predictive models to help sales personnel who manage hundreds of stores better predict what stores they should focus on. The team has also built models for the Composites division using inventory and railcar tracking data to help them predict if inventory levels of raw materials will dip too low. These raw materials are fed into their furnaces, 24/7. If the materials do not get to the furnace on time, the company incurs cost. The CoE has also built models for marketing teams that use weather data and census data to help predict the impact storms have on product sales.

Analytics development and deployment processes are key to success, according to Burns. "We utilize an analytics playbook that includes the charter for each product, signoffs, technical design—everything from beginning to end. You can't move a model into production unless testing is validated. All of this—the blueprint, the testing reviews—is in the playbook."

This playbook is part of a structured project management function that uses a portfolio process. The playbook methodology, along with good project management, has helped Owens Corning reduce issues such as scope creep and changing requirements. This helps Owens Corning be faster and more efficient.

Another best practice Burns cites is training. A couple of years ago, his team gamified the training program. Originally, they started by showing employees how to query and then retrieve the data. By the time they got to how to use the data for insights, most employees were checked out and overwhelmed. They could not get past the database and data modeling. "We decided to turn our training method upside down," said Burns. "Now we show them how to get insights from data. If they want to go deeper, then we train them on the database and modeling tools."

The company also holds an Analytics Challenge competition. The team selects finalists to showcase what they have done. "These competitions have given leadership visibility into the talent that exists at the company and shows what is possible with analytics," said Burns. This generates more ideas and opportunities to pursue with analytics.

The Future of AI

There is much to say about AI. In this report, we've primarily focused on best practices for AI in business. However, there are many opportunities and challenges for AI in the next three to five years and beyond. We wanted to get input from respondents about what they believe to be some of the top challenges for AI in the near future, as well as what they think is most promising in the field of AI (Figure 12).

- **Access to talent is a top challenge.** As Figure 12 illustrates, access to talent to build AI applications was cited by 50% of respondents. Skills, as we've seen already in this report, are a top concern. Skill gaps include architects to build the complex infrastructure that may be required for AI, data engineers to build data pipelines, data scientists to build models, DevOps to put models into production, and developers to build AI apps. Depending on the apps, other skills might be needed as well, such as computational linguists to deal with natural language and ontologies. Scientists and other subject matter experts would be needed for apps that rely on complex, in-depth knowledge.

  As one interviewee said, "If you're looking at biological pathways for Parkinson's disease, you might think that the ML results look interesting. However, if these results aren't defined in a Parkinson's disease ontology, then it's either a completely radical new discovery and you need to modify your ontology or, more likely, the ML gave you odd correlations." Vendors are making tools easier to use, but as discussed in this report, that doesn't entirely eliminate the need for trained staff.

Explainability is going to be critical for AI in the next few years.

- **The ability to explain outcomes is also key.** Close behind access to talent was the ability for AI to explain outcomes. Forty-eight percent of respondents cited this as a challenge. This ties into the use of automated tooling, such as AutoML and other tools. Respondents recognize that it will be important for these tools to be transparent.

  New privacy rules will also require explainability. For example, the GDPR requires that any analytics outcome be explainable to a consumer who asks about it. The California Consumer Privacy Act (CCPA) will also require explainability. That means that even if model building can be automated, someone needs to understand the model. We find it interesting that over a third (38%) of respondents cited privacy issues associated with AI as a challenge. Thirty-two percent cited regulatory issues.

- **Integrating data for AI.** Another area that respondents see as challenging is integrating data for AI. This was discussed earlier in the infrastructure section. Data for AI will be coming from multiple sources across a multiplatform environment. Integrating data across platforms will become key. Organizations will need to find good tooling to help with this.

- **Ethics and displacing workers are not major concerns.** Respondents were not greatly interested in some of the concerns that have been raised about AI. For instance, though 30% of respondents did cite the importance of developing ethical applications for AI, only 11% cited AI displacing workers as a top concern.

What top challenges do you see for AI in the next 3-5 years. Please select no more than three responses.

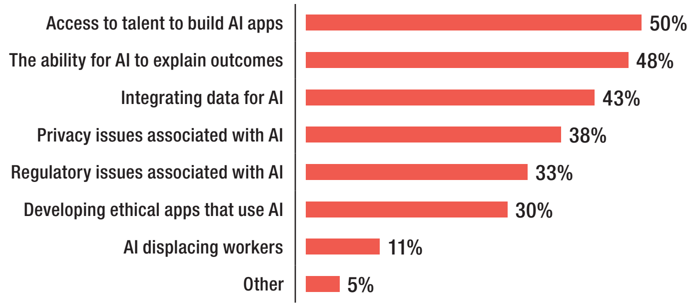

Figure 12. *Based on 256 respondents from both the active and investigating group. A maximum of three responses allowed.*

We also asked respondents what they see as the most promising aspects of AI (Figure 13).

- **Insights ranked at the top of the list.** More than one-third (38%) of respondents cited better insights as the most promising aspect of AI. Along with this, 26% cited the ability to automate decisions. This is a chief value proposition of the technology.
- **AI for good ranked low on the list.** On the opportunities front, there are many initiatives underway for AI for good. Vendors have started to deploy their own programs. They are working with nonprofit groups and others, looking at ways to use AI for everything from cancer research to helping endangered species. A number of startups have also emerged. Even so, AI for good ranked at the bottom of the list, on par with the idea of general intelligence.
- **Robotic processes and AI for menial tasks.** In the middle of the list, using AI for menial tasks and RPA garnered about 10-13% each. Although we've seen in the research that RPA can provide a lot of value, this wasn't cited as the most promising aspect of AI.

What do you see as the most promising aspect of AI?

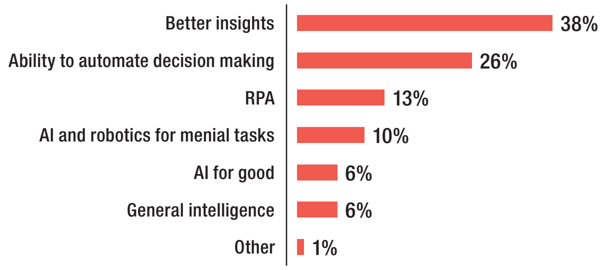

Figure 13. *Based on 256 respondents from both the active and the investigating group.*

Recommendations

This report has detailed many best practices for AI. In closing, we summarize the report by listing the top best practices for successful AI and digital transformation, along with a few comments about why each is important. Think of the best practices as recommendations that can guide your organization into successful model implementations.

Know your business problem. It is important to start with a *real* business problem and a clear objective when embarking on a more advanced analytics project. The same holds true for AI. Organizations that had experience with AI and had succeeded cited having a concrete business objective in mind. This ties into a second best practice which is to find basic, achievable projects where AI can help improve time and cost objectives. "Going big or going home" ranked at the bottom of the list of AI best practices. Be realistic with which projects will have an impact and build on your successes.

Get executives on board. We have seen in this report that organizations deploying AI often have CEO support. It can be easier to get AI funding if you can get an executive to sponsor your effort. The executive can also evangelize the results, use the results, and help build trust in AI. This can go a long way in helping AI to become more widespread in your company and in getting people on board.

If your organization wants to do sophisticated analytics, it is going to have to hire data scientists, data engineers, and DevOps.

Plan for new skills. There are excellent tools on the market that can help everyone become more productive. However, the reality is that if your organization really wants to conduct sophisticated analytics or build AI apps, you're going to have to hire data scientists along with other skill sets such as data engineers, developers, and DevOps (and others, depending on the business problem). If you're planning to improve the skills of existing talent for any of these roles, make sure to set aside funding for training.

Consider open source and commercial tools for building AI models and analytics. Open source tools such as R and Python have become go-to standards for building models and innovating with analytics. Part of this is because data scientists are often trained in open source tools. However, that doesn't mean that commercial tooling is going away. Many data scientists use a range of tools, including open source and commercial tooling. Companies that want to keep their data scientists from leaving will support both. Additionally, commercial tooling often provides functionality that open source does not, such as model registries and other features needed for implementing AI in production.

Embrace new data types. "Newer" data types can help to drive insights and innovation and are important for digital transformation. Consider moving past using only structured data for AI initiatives. The active group is already doing this; 85% of the active group is using text data; over half are using image, IoT, and streaming data.

Consider modern data management tools. AI may require new data types. Additionally, there is a good chance that your data will be distributed across multiple platforms. That means that your organization will need to consider modern tools to deal with modern data. A mistake that organizations make is that they hire a few data scientists but don't invest in the data infrastructure. Data integration will be critical. Tools that can help integrate data across platforms (including the cloud) will be needed. This includes modern data cataloging tools that can help users access, understand, and trust their data.

Think about modern pipelining tools. Along with modern data management tools, it is important to plan for new pipeline tools. This may include data pipelining tools that infuse ML to help automate some data integration and preparation steps. It may also include completely different kinds of tooling to help extract useful data from text. That includes text mining or text analytics tools. Traditional and new vendors offer these tools. Some of vendors offer the tools as part of an analytics platform.

Build apps! Once you get your AI efforts off the ground, consider building applications to monetize your AI efforts. As we saw in this report, the active group was more likely to think that apps are critical for moving forward to compete. This should be a heads up to those investigating the technology now.

Plan for production. Many organizations make the mistake of not planning for AI in production. That means that they don't have a good way of validating models, putting them into production, tracking them, monitoring them, and updating them when they go stale. This is the nuts and bolts of digital transformation. Not surprisingly, the active group cites deploying models into production as a bigger challenge than the investigating group. Successful organizations often have a DevOps team responsible for model deployments. This will be crucial as organizations roll out their ML models and other AI solutions.

Stay abreast of new technologies. New technologies for AI are entering the market rapidly. These include new services sold in vendor marketplaces that can be used for insights and apps, new open source tooling, and new technologies engineered for big data and distributed architectures. Keep up with the changes, especially if your company wants to stay competitive. The active group was more likely to say that AI is disrupting their industries compared to the investigating group. No doubt, the active group is staying educated about how these technologies can help.

Don't forget about the people part of the equation. Digital transformation with AI requires change. That change involves technological as well as cultural change. As we saw, AI will require new roles. It will also require that people buy into the projects and trust the output. It may also change how people do their jobs and may create new jobs (e.g., more DevOps, app developers, or those who deal with new product management). It will be important to plan for these changes.

answerrocket.com

An AI-Driven Analyst for Impact Makers

With the ever-increasing wealth of data now available to inform smarter decision making, companies need a better way to democratize access to data, analytics, and, ultimately, insights. Business users want direct access to vital data sources to ask questions and get answers on demand. Data and analytics professionals want to ensure the source data, the analytical approach, and the resulting output are sound, consistent, and accurate.

AnswerRocket was founded in 2013 on the idea that data and analytics should be accessible to all who need them. With AnswerRocket, data exploration, analysis, and insight generation are adapted to the user rather than requiring the user to adapt to yet another new technology. Our mission is to eliminate the drudgery of human analysis through intelligent automation.

From the start, we've leveraged AI and machine learning technology to unleash the power of augmented analytics across organizations. The results are faster insights, better decisions, and more time to dedicate to executing ideas versus churning through data analysis.

Democratize Analytics with AnswerRocket

AnswerRocket's AI-driven analytics platform leverages augmented intelligence to help businesspeople make faster insight-driven decisions.

Conversational Search
With conversational search and the power of natural language processing, business people can ask questions in plain English and get quick answers and insights on demand.

Intelligent Analysis
AnswerRocket leverages AI and machine learning techniques to thoroughly analyze your data with automation, going beyond just business intelligence and descriptive analytics to also handle diagnostic and predictive analysis.

Interactive Visualizations
AnswerRocket produces answers in the form of optimized, interactive visualizations. Our smart platform automatically determines the most appropriate chart or graph type to display your answer.

Smart Insights
Leveraging natural language generation—AI technology that translates data into human language—AnswerRocket is able to compose explanatory narratives and valuable insights that accompany analysis.

RocketBots
Unlock data science capabilities with AnswerRocket's pioneering RocketBots. These AI agents automate time-intensive analytics workflows, invoking advanced machine learning capabilities to run complex, high-value analysis in a matter of seconds.

Loved by Data & Analytics Teams

Designed to deliver powerful BI and analytics capabilities anywhere, anytime, AnswerRocket works seamlessly with a company's existing data infrastructure and enterprise requirements. There's no need to move your data and no need to invest in costly hardware.

A fully extensible platform allows data science teams to easily operationalize machine learning models, supported by enterprise-grade scalability, administration, security, and governance. AnswerRocket's toolkit also enables data scientists to build and deploy new algorithms with data science notebooks such as Jupyter and top AI frameworks and machine learning libraries such as scikit-learn.

Ultimately, AnswerRocket helps companies solve problems and achieve business outcomes. Through machine learning, augmented intelligence, and open source AI, AnswerRocket drives digital transformation for both data and business personnel.

research

TDWI Research provides research and advice for data professionals worldwide. TDWI Research focuses exclusively on data management and analytics issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of data management and analytics solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

tdwi

Transforming Data With Intelligence™

555 S. Renton Village Place, Ste. 700
Renton, WA 98057-3295

T 425.277.9126
F 425.687.2842
E info@tdwi.org

tdwi.org